Abnormal Behavior Detection using Machine Learning in a Virtual Mobile Cloud Infrastructure

[1]Naren Raghavendra Suri, [2]S.Gowtham Bharath,

Abstract: —
The availability of technology to deliver services that human relied today has become ease through mobile devices. With the increased demand for mobiles and smart phones, the mobile service providers are changing their service architecture. In order to cope up with the demand and necessity of services, the mobile cloud infrastructure has become very vital. In this paper we explore the detection of abnormal behavior of users in new mobile cloud infrastructure that forms as a bridge between mobiles and cloud services. The present cloud infrastructure would need a better means to control the abnormal behavior over mobile devices, which is explored and explained using various sample cases those leads to that security lapse, which is detailedusing a methodology to detect abnormal behavior on mobile cloud test best using the Random Forest technique. Random Forest is a well-known technique to detect abnormal behavior, we are applying the same technique to the Virtual mobile cloud infrastructure to detect and classify users on the instance.

Index Terms—Random Forest,Decision tree, abnormalbehavior detection, virtual cloud services,Secure communication on virtual mobile cloud infrastructure.

I. INTRODUCTION

The diverse electronic mobile phones released every day have been increasing in their number. Statistics says that by 2012 there are almost 3 million android and 4 million iPhone's in the market.

The technological improvements had got various features like phonebook, enriched call, push notification, and content sharing of multi-media. These mobile features are provided through the cloud infrastructure to provide an enhanced computational speed and service on demand.

The data of the users were moved to cloud environment and this data can be used from anywhere through the mobile devices. We have taken the virtual mobile instances on the cloud environment to understand the security to be provided to them. There are certain apps in the market which can help us to understand the miss-use of a device. However, having such applications on each and every device and updating them frequently. Making users to understand the bits and pieces related to their security would make using the technology difficult for them. Moreover, the burden on the devices would be high as well. In order to avoid such practical issues we have implemented the idea of observing the human behavior with the operations they do on the cloud environment. With that we would like to address the issue of security at the high level rather than asking every instance of mobile to have the signature based solutions.To understand the behaviors of users, we have created a test bed environment with the various mobile instances. These mobile instances are monitored for both host and network data to predict the abnormality. The malicious actions/ programs are run to test the abnormality of users on the test bed environment.

Service Scenarios on Test Bed:

We have different use case scenarios of the people who will be using the virtual mobile infrastructure. For instance we have categorized users in to three categories. The general user, Developer, and Advanced user. Each user may consume content of different size through different place and different device.

According to the present scenario, The General user is the one who uses general basic services like texting etc. Most of their activities are fixed, however there will be certain randomness.

The Developer: These users use the environment extensively and they have a different behavior in the classification. Since these users are very aware of all the activities that occur in the cloud environment their activities vary very much.

Advanced Users: These users are in between to both the general user and a developer. They have a fair understanding of the system and the actions they perform would resemble a active and flexible environment. All these three classifications would be

helpful to distinguish the behavior of each group. However on the abnormal behavior, they are caught and resembled by the Machine learning algorithm that we are using.
One instance of user nature and content he/she may consume: [1]
Type of user: Advanced User
Place of access: Office
Device: smart phone
Network: office Wi-Fi
Consuming content: smart work

We have taken few such example scenarios to illustrate the user groups for the classification, and to define the user behavior.

Random Forest technique is applied to classify the users and their behaviors to detect the abnormal user actions on the test bed. The data collected with various features on the actions that can be performed by each user type is stored. The complete data collected through the cloud test bed is taken as a vector.
V =(x1, x2,x3,……, xd). The random forest technique operates on the decision made at the each node. The given data set is taken as V, and the subset of the data is taken and a decision tree is computed for different random data sets as well as the different features.
$S_{j=} {}^{L}{}_{j} U S^{R}_{j}$

The information gain is computed to decide the classification that can give the best results on classification.

$I_i = H(S_j) - \sum_{i} ((|S^i_j| / |S_j|)H(S^i_j))$
i∈ {L,R}

The Algorithm Implemented: [3]

1. For a= 1 to A
   (a) Select a sample Q* of size N from the training data.
   (b) Grow a random forest tree T_a to be bootstrapped data, by recursively repeating the following steps for each terminal node of the tree, until the minimum node size $n_{min\ is}$ reached.
      1. Select m variables at random from the p variables.
      2. Pick the best variable/split-point among the m.
      3. Split the node in to two daughter nodes.

2. Output the ensemble of trees $\{T_a\}^{A}_{1}$

In order to illustrate the activity that was done exactly on test bed, we would like to illustrate an example here. We have collected the data and the decade long features to analyze the behavior of the users. The most important findings were done under three categories,

1. Type 1: In this the users are not visible but the data of a mobile instance is transferred to the external server or user.
2. Type 2: Here the mobile instance itself gets spoiled as zombie and uses the instances to construct botnet and DDos attacks.
3. Type 3: it maximizes the use of the network unnecessarily and keeps the network chocked.

Abnormal Behavior Detection:

We have used the RF algorithm mentioned earlier to handle the random forest approach for our analysis.
The collected sample data is classified and used to detect the abnormal behavior at the later stages. We have described three states in understanding the behavior of the cloud instance users. The three categories are, inactive, active and abnormal. Inactive users don't avail the services as name specifies. Active users access the cloud instances with no abnormality detected. Abnormal users, the behavior of these users stand suspicious. From the virtual mobile host we collected the data and after applying the algorithm to classify the abnormal users, we have got 56 inactive users 146 active users and 45 abnormal activities by users on the virtual instances.

Related Work:

The previous methods used to observe the nature of the mobile devices by tracking the applications and their behavior very much asanother spamming issue. This procedure was not really efficient and could not reveal much abnormal behavior. We the randomness of the Random Forest approach, we are able to detect many abnormal behaviors on a cloud infrastructure which makes this a better option.

Conclusion:
With the help of RF algorithm we can classify the data using and detect the abnormal behavior of the users. The monitoring architecture that we proposed would help to monitor actions directly on the cloud instance. In addition to this we would like to explore the

detection of various behaviors to a higher extent by collecting various other types of malwares. We would also like to increase the efficiency while increasing the features and types of malwares.

REFERENCES

[1] The Gnutella protocol specification, 2000. http://dss.clip2.com/GnutellaProtocol04.pdf.
[2] The Algorithm is from the Book of Hastie, Friedman and Tibshirani
[3] Cubrilovic, N. "Letting Data die a natural death", International Journal of electronic Government Research, 2009.
[4] Manish Pokharel and Jong Sou Park. "Cloud computing future solution for e-Governance", Proceedings of 3rd International Conference on Theory and Practice of Electronic Governance, IEEE 2009.
[5] Google Docs Privacy Policy (Version of 3 Octobe2010). <http://www.google.com/intl/en/privacypolicy.html>, at 10 October 2010.

[6] I. Clarke, O. Sandberg, B. Wiley, and T. W. Hong. Freenet: A distributed anonymous information storage and retrieval system. ICSI, Berkeley, CA, USA

[7] Dan Svantesson, Roger Clarke. "Privacy and consumer risks in cloud computing". Privacy consumer risks journal, pages 391-397, July, 2010.

[8] Frank Gens. "IT Cloud Services User Survey, part 2: Top Benefits and Challenges", Survey conducted by IDC, October, 2008.

[9] Buyya R, Yeo, Venugopal CS, S Broberg, J Brandic, I. "Cloud Computing and Emerging IT Platforms: Vision, Hype, and Reality for Delivering Computing as the 5th Utility". Future Generation Computer Systems 25, pages 599-616, 2009.

[10] J.D Blower. "GIS in the cloud: Implementing a web map service on Google App Engine", Proceedings of the 1st International Conference and Exhibition on Computing for Geospatial Research and Application, Washington D.C, June 21-23, 2010.

[11] Mark Nicolett, Jay Heiser. "Accessing the security risks of cloud computing", Gartner Inc., June, 2008.

[12] W. J. Bolosky, J. R. Douceur, D. Ely, and M. Theimer. Feasibility of a serverless distributed file system deployed on an existing set of desktop pcs. In

Proc. ACM SIGMET- RICS'2000 , pages 34–43, 2000.
[13] Ortuatay B. "Twitter service restored after hacker attack", Journal of the Baltimore Sun, 2009.
[14] Scott Paquette, Paul T.Jaegar, Susan C.Wilson. Identifying the security risks associated with governmental use of cloud

First Author: ***Naren Raghavendra Suri***received B.Tech inInformation Technology, in the year 2010. He is currently working for Computer Sciences Corporation from last 4 years. And he is interested in the field of Machine Learning - Artificial Intelligence, Cloud Computing and Network Security.

Second Author: ***GowthamBharathSrugarapu***receivedB.Tech in Computer Science and Engineering He is currently working for Computer Sciences Corporation from last 6 years. And he is interested in the field of Cloud Computing and Network Security.